我所理解的爬蟲流程?

- •技術基礎: urllib、requests、 selenium
- 進階技術: 多執行緒抓取、登入抓取、 突破ip限制和伺服器抓取

獲取網頁

我要請求囉

- •尋找消失的API/URL
- •溝通方式: OPTIONS / GET/HEAD/POST/PUT/DELETE/TRACE/CONNECT
- •常用功具:

程式請求: urllib3/requests

模擬請求: selenium

- •工具語言:
- BeautifulSoup/PyQuery/ JSON/Scrapy
- •CSS選擇器: Selenium
- •路徑語言: Selnium/Ixml

優化程式

- 模組化程式設計
- 定期定時的爬取觸發機制
- 異常處理設計

解析網頁